

## APPARATUS AND METHOD FOR PROBE DESIGN

Publication number: JP2002345480 (A)

Publication date: 2002-12-03

Inventor(s): MATSUMOTO TOSHIKO; NAKASHIGE AKIRA; NOZAKI YASUYUKI; UENO SHINGO; TAMURA TAKURO +

Applicant(s): HITACHI SOFTWARE ENG +

Classification:

- international: C12M1/00; C12N15/09; C12Q1/68; G01N1/00; G01N33/53; G01N33/566; G06F19/00; C12M1/00; C12N15/09; C12Q1/68; G01N1/00; G01N33/53; G01N33/566; G06F19/00; (IPC1-7): C12M1/00; C12N15/09; C12Q1/68; G01N33/53; G01N33/56

- European: G06F19/20

Application number: JP20010161034 20010529

Priority number(s): JP20010161034 20010529

Also published as:

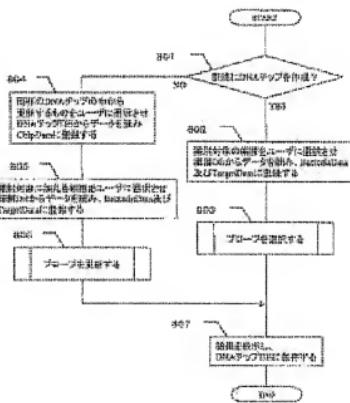
JP3979793 (B2)

US2003069701 (A1)

US7035738 (B2)

## Abstract of JP 2002345480 (A)

PROBLEM TO BE SOLVED: To reduce time and money costs in the case of producing a DNA chip corresponding to a bacterium as a discrimination object added afterwards and to maximize discrimination precision of DNA contained in a specimen even in a case wherein a unique probe can not be designed in a DNA sequence as a discrimination object. SOLUTION: When a probe is designed at the beginning, a plurality of different probes are prepared for one kind of a bacterium. When a bacterium as a discrimination object is newly added and a part of the probes is not used, the residual probes are used and discrimination is carried out. In a case wherein a unique probe can not be designed in the bacterium as the discrimination object, a plurality of probes are used to increase the probability of right discrimination.; A set of probes for maximizing the probability of the right discrimination is selected by taking possibility of a plurality of kinds of bacteria existing at the same time into consideration.

Data supplied from the **espacenet** database — Worldwide

**Family list**

2 application(s) for: JP2002345480 (A)

Sorting criteria: Priority Date Inventor Applicant Ecla

**1 APPARATUS AND METHOD FOR PROBE DESIGN****Inventor:** MATSUMOTO TOSHIKO ;  
NAKASHIGE AKIRA (+3)**EC:** G06F19/20**Publication** JP2002345480 (A) - 2002-12-03**Info:** JP3979793 (B2) - 2007-09-19**Applicant:** HITACHI SOFTWARE ENG**IPC:** C12M1/00; C12N15/09; C12Q1/68; (+16)**Priority Date:** 2001-05-29**2 Probe designing apparatus and probe designing method****Inventor:** MATSUMOTO TOSHIKO [JP] ;  
NAKASHIGE RYO [JP] (+3)**EC:** G06F19/20**Publication** US2003069701 (A1) - 2003-04-10**Info:** US7035738 (B2) - 2006-04-25**Applicant:** MATSUMOTO TOSHIKO, ;  
NAKASHIGE RYO, (+4)**IPC:** C12M1/00; C12N15/09; C12Q1/68; (+15)**Priority Date:** 2001-05-29Data supplied from the **espacenet** database — Worldwide

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-345480

(P2002-345480A)

(43)公開日 平成14年12月3日 (2002.12.3)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーラード <sup>8</sup> (参考)
C 1 2 N 15/09	ZNA	C 1 2 M 1/00	A 4 B 0 2 4
C 1 2 M 1/00		C 1 2 Q 1/68	A 4 B 0 2 9
C 1 2 Q 1/68		G 0 1 N 33/53	M 4 B 0 6 3
G 0 1 N 33/53		33/566	
33/566		C 1 2 N 15/00	ZNA
			審査請求 未請求 請求項の数10 O L (全 16 頁)

(21)出願番号 特願2001-161034(P2001-161034)

(71)出願人 000233055

日立ソフトウェアエンジニアリング株式会社

(22)出願日 平成13年5月29日(2001.5.29)

神奈川県横浜市中区尾上町6丁目31番地

(72)発明者 松本 俊子

神奈川県横浜市中区尾上町6丁目31番地  
日立ソフトウェアエンジニアリング株式会社内

(74)代理人 100091096

弁理士 平木 勉輔 (外1名)

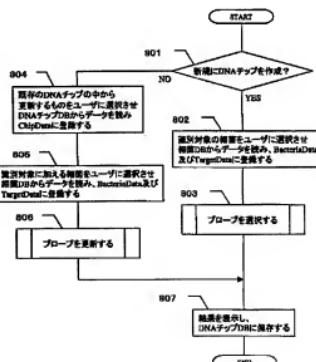
最終頁に続く

## (54)【発明の名称】 プローブ設計装置及びプローブ設計方法

## (57)【要約】

【課題】 後から加わった識別対象の細菌に対応したDNAチップを製造する場合の時間的・金銭的コストを削減し、また、識別対象のDNA配列にユニークなプローブが設計できない場合にも試料中に含まれるDNAの識別精度を最大化する。

【解決手段】 最初にプローブを設計するときに、一種類の細菌に対して複数の異なるプローブを用意しておく。そして、新たに識別対象の細菌が加わり一部のプローブが使えなくなった場合には、残りのプローブを用いて識別する。識別対象の細菌にユニークなプローブが設計できない場合には、複数のプローブを用いて正しく識別できる確率を高くする。また、複数種類の細菌が同時に存在する可能性を考慮して、正しく識別できる確率を最大化するプローブの組を選別する。



## 【特許請求の範囲】

【請求項1】 ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計装置において、塩基配列を含む識別対象ターゲットの情報を登録する識別対象ターゲット登録処理部と、登録された各ターゲットに対してその塩基配列の部分配列から各ターゲットにユニークな部分配列をプローブとして選択するプローブ選択処理部と、前記プローブ選択処理部においてユニークな部分配列が選択できなかったターゲットがある場合、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択する複数プローブ代用処理部とを含むことを特徴とするプローブ設計装置。

【請求項2】 ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計装置において、複数種類のターゲットを識別するため既存のプローブセットの情報を読み込む手段と、識別対象に追加する新たなターゲットの塩基配列を含む情報を読み込む識別対象ターゲット追加処理部と、前記新たなターゲットの追加によって前記既存のプローブセット内のプローブによる識別が困難になったターゲットを検出するプローブ再検討処理部と、前記新たなターゲットを識別するためのプローブ及び前記プローブ再検討処理部によって検出されたターゲットを識別するためのプローブを追加するプローブ追加処理部と、

前記プローブ追加処理部においてユニークなプローブが選択できなかったターゲットがある場合、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択する複数プローブ代用処理部と、

前記プローブ追加処理部及び/又は複数プローブ代用処理部による処理結果を反映してプローブセットを更新するプローブ更新処理部とを含むことを特徴とするプローブ設計装置。

【請求項3】 請求項1又は2記載のプローブ設計装置において、前記複数プローブ代用処理部は、複数のプローブとハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を一つの基準とし、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計装置。

【請求項4】 請求項1又は2記載のプローブ設計装置において、前記複数プローブ代用処理部は、混合感染・合併症の情報を用いて複数のプローブとハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を評価し、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計装置。

【請求項5】 ハイブリダイゼーションを介して識別対

象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計方法において、識別対象であるターゲット集合の塩基配列を含むデータを読み込むステップと、

読み込まれた各ターゲットの塩基配列の部分配列から各ターゲットにユニークな部分配列をプローブとして選択するステップと、

前記ステップにおいてユニークな部分配列が選択できなかったターゲットがある場合、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択するステップとを含むことを特徴とするプローブ設計方法。

【請求項6】 ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計方法において、複数種類のターゲットを識別するため既存のプローブセットのデータを読み込むステップと、識別対象に追加する新たなターゲットの塩基配列を含むデータを読み込むステップと、前記新たなターゲットの追加によって前記既存のプローブセット内のプローブによる識別が困難になったターゲットを検出するステップと、

前記新たなターゲットを識別するためのプローブ及び前記識別が困難になったターゲットを識別するためのプローブを追加するプローブ追加ステップと、

前記プローブ追加ステップにおいてユニークなプローブが選択できなかったターゲットがあったとき、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択する代用プローブ選択ステップと、前記プローブ追加ステップで追加されたプローブ及び/又は前記代用プローブ選択ステップによって選択されたプローブを反映してプローブセットを更新するステップを含むことを特徴とするプローブ設計方法。

【請求項7】 請求項5又は6記載のプローブ設計方法において、前記代用プローブ選択ステップは、複数のプローブとハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を一つの基準とし、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計方法。

【請求項8】 請求項5又は6記載のプローブ設計装置において、前記代用プローブ選択ステップは、混合感染・合併症の情報を用いて複数のプローブとハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を評価し、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計方法。

【請求項9】 請求項5～8のいずれか1項記載のプローブ設計方法をコンピュータに実行させるためのプログラム。

【請求項10】 請求項5～8のいずれか1項記載のア

ロープ設計方法をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ターゲット中に含まれる複数種類のDNAを判別することを目的とするDNAチップの技術に関する。

【0002】

【従来の技術】近年の遺伝子解析技術の発展によって、遺伝子の機能や構造が次第に明らかになってきた。中でもDNAチップあるいはDNAマイクロアレイ（以下、まとめてDNAチップという）の技術は、遺伝子解析に有効な手段であるとして注目されている。

【0003】DNAチップとは、ガラス、シリコン、プラスチックなどの基板の表面上に多数の異なったDNA（プローブ）が高密度に配置されたものである。プローブとしては、通常DNA（PCR断片）や、20~30mer程度の短鎖ヌクレオチドなどが用いられる。DNAチップの原理は、DNAを構成する4つの塩基A（アデニン）、T（チミン）、G（グアニン）、C（シトシン）において、AとT、GとCが互いに水素結合する性質、すなわちハイブリダイゼーションに基づいている。蛍光物質などで標識したDNAあるいはRNA（ターゲット）は、このDNAチップ上のプローブとハイブリダイゼーションすることで捕獲される。捕獲されたDNAは各スポットからの蛍光シグナルとして検出され、これをコンピュータでデータ解析することにより試料中の数千から数万のDNAの状況を一挙に観測できる。

【0004】DNAチップの利用方法の一つとして、試料中の遺伝子（またはDNA断片）を捕獲することで、調べたいDNAが試料に含まれているか観測したり、捕獲したDNAの配列を読み取ったり、SNPなどのDNAの多型部分を調べる利用法（SBS法）がある。DNAチップの利用法の一例として、臨床検査における細菌同定を挙げて説明する。従来、臨床検査で細菌を同定するためには、細菌を培養して形状や生化学的性質を観察したり、免疫反応を確認したりしなくてはならなかったため、多くの日数がかかる。また、複数種類の細菌を同定するためには、それぞれの種類について別々に生物学的実験を行う必要があった。そのため、臨床の現場では時間やコストの問題から、分離頻度の高い細菌だけを対象として同定を行っており、分離頻度は低いが重篤な症状を呈する細菌の同定精度が悪いという問題点があった。

【0005】DNAチップを用いた細菌同定では、DNA配列データベースなどから各細菌のDNA配列情報を入手する。細菌毎にDNA配列から特異的な部分配列を選択し、これをプローブとしてDNAチップ上に配置する。一方、患者から採取した血液や痰などの試料から細菌のDNAを抽出し、PCRで増幅してDNAチップと反応させる。

【0006】図1はこれを模式的に表した図である。細菌、Q、Rなど20種類の細菌についてDNA配列からそれぞれ特異的な部分配列を選択し、DNAチップ上に配置している。これらのプローブの位置は横No（1~5）と縦No（1~4）で表現する。患者の血液・痰などから抽出された細菌のDNAを含む試料110とDNAチップ100上のプローブとを反応させた結果、（横No1、縦No2）と（横No3、縦No5）に対応する二つのスポットからシグナルが観測されたとする。この時、プローブの位置情報を管理する表から、細菌(*Actinobacillus actinomycetemcomitans*)（横No1、縦No2）と細菌(*Klebsiella oxytoca*)（横No3、縦No5）が試料に混入していると分かること。

【0007】このように、DNAチップを用いた細菌同定では、チップ上に複数種類のプローブを配置できるため、一枚のチップで多くの種類の細菌を網羅的に検出できる。また、細菌を培養する必要がないため、検出に要する時間も短縮できる。DNAチップを用いた細菌同定は将来、病院、保健所、検疫機関などで利用可能になると期待されている。これらの機関では、DNAチップに配置するプローブを設計したりプローブやDNAチップ作成装置を用意したりする必要はなく、購入したチップを用いて検出すればよいので、簡便に利用できる。

【0008】

【発明が解決しようとする課題】実際の現場でこのようなDNAチップを利用するためには、二つの課題がある。一つ目の課題は、DNAチップ上に配置するプローブを設計した後で新たに識別対象のDNAが加わった場合に、新たな細菌に対応したDNAチップを製造するために多くの手間を要する点である。例えば、臨床検査や食品検査などにおける実際の細菌同定では、下の例に挙げるようになにまで目されていなかった細菌や突然変異で現れた新種の細菌が流行し、検出・識別の必要が生じる場合がある。

【0009】例1：突然変異により薬剤耐性を獲得した黄色ブドウ球菌や結核菌が出現した

例2：食中毒が流行したので、原因となった細菌の検出需要が急増する

このような新しい細菌は、それまで用いていたプローブでは検出・識別できない。したがって、新たに加わった細菌にも対応できるよう、DNAチップ上にスポットするプローブの情報を更新し、新しいプローブ情報に基づいて製造されたDNAチップを用いて検出を行う必要がある。

【0010】しかし、新たに識別対象の細菌が加わることにより、既存の細菌に対応するプローブも変更しなくてはならない。現在提案されているDNAチップ用プローブ設計方法は、まず、それぞれのDNA配列に対してユニークな部分配列を列挙してプローブの候補とし、次に、それぞれのスポットでの反応温度が均一になるように

する

- ・プローブが自分自身で絡み合わない
- ・誤ってハイブリダイゼーションしてしまう確率を小さくする

などの条件を考慮し、候補を絞り込んで最適なプローブの組合せを得る (Ken-ichi Kurata et al.: Probe Design for DNA Chips: Genome Informatics 1999, 225-6, 1999)。ところが新たに識別対象の細菌が加わると、プローブがユニークでなくなったり、反応温度が均一でなくなったり、誤ってハイブリダイゼーションする確率が大きくなったりするため、多くのプローブが最適でなくなる恐れがある。

【0011】図2は、新たに識別対象の細菌が加わったことによりプローブが最もでなくなる様子を示したものである。実線で囲まれた枠円は、DNAチップ新規設計時、すなわち、識別対象の細菌が加わる前の、プローブ選別の過程を示す。それぞれの枠円は、部分配列全体200、DNAチップ設計時ユニークな部分配列の集合211、設計時ユニークかつ反応温度が均一になるプローブ候補の集合212、設計時ユニークかつ反応温度が均一かつ自分自身で絡み合わないプローブ候補の集合213、設計時ユニークかつ反応温度が均一かつ自分自身で絡み合わないかつ誤ってハイブリダイゼーションする確率が低いプローブ候補の集合214を表す。図示するプローブ候補221、222、223はそれぞれ、反応温度条件を満たさない、自分自身で絡み合う、誤ってハイブリダイゼーションする確率が高い、という理由で候補から除外される例である。このように条件を強めていてプローブ候補を絞り込み、一番内側の集合214に含まれるプローブ231、232を用いていた。

【0012】その後、識別対象の細菌を追加し、それぞれの条件を満たす集合が実線から点線へ変わってしまったとする。すなわち、点線で示す集合241は細菌追加時ユニークな部分配列の集合、集合242は細菌追加時ユニークかつ反応温度が均一になるプローブ候補の集合、集合243は細菌追加時ユニークかつ反応温度が均一かつ自分自身で絡み合わないプローブ候補の集合、集合244は細菌追加時ユニークかつ反応温度が均一かつ自分自身で絡み合わないかつ誤ってハイブリダイゼーションする確率が低いプローブ候補の集合である。その結果、プローブ231は誤ってハイブリダイゼーションする確率が高くなってしまい、プローブ232はユニークでなくなるので、これら2つのプローブは、新たに細菌が加わったことにより最適でなくなってしまう。

【0013】新しく識別対象に加わった細菌に対応するため、新しい細菌を含む入力に対して上記アルゴリズムを適用して最適なプローブの組合せを選び直すと、図2に示したプローブ231やプローブ232のようなく多くのプローブを差し替えなくてはならなくなる。

【0014】DNAチップ製造会社では、新たな細菌に対

応したDNAチップを製造するために、新しく追加されたプローブ毎に検証実験を行ってその有効性を確認し、新しいプローブ用DNA断片を用意する必要があるため、多くの時間とコストを要する。また、DNAチップを用いて細菌検出を行う病院、保健所、検疫機関などでは、今までに購入したDNAチップに代えて新たな細菌に対応したDNAチップを購入して細菌検出を行わなくてはならない。特に、即時性が要求されている現場では、新たなDNAチップを購入するまで長い時間がかかるのは大きな問題である。

【0015】二つ目の課題は、必ずしもそれぞれの識別対象のDNAにユニークなプローブが選択できるとは限らない点である。プローブがユニークであるとは、そのDNA配列にはプローブが部分配列として含まれるが、他のDNA配列には部分配列として含まれないことである。ユニークなプローブを用いれば、他のDNAの有無に関わらず「このプローブからシグナルが観測されればこのDNAが混入している。観測されなければ混入していない」という判断ができる。それぞれの種類のDNAについてユニークな部分配列をプローブとして用意しておけば、全てのDNAの組合せを正しく判断できる。

【0016】しかし、識別対象のDNAの一つが他のDNAと非常に似通っており、ユニークな部分配列が選択できないことがある。図3は、DNA配列にユニークな部分配列が存在しない例である。RのDNA配列は、PやQと共通であるため、ユニークな部分配列が選択できない。上記のプローブ設計方法は、それぞれのDNA配列に対してユニークな部分配列が存在すると仮定しているため、このようないくつかのDNAには対応できない。

【0017】本発明では、従来技術の問題点に鑑み、識別対象のDNAの種類が後から加えられたときに柔軟に対応できるプローブ設計法を提供することを目的とする。また、識別対象のDNA配列にユニークなプローブが設計できない場合にも、試料中に含まれるDNAの識別精度を最大化するプローブ設計法を提供することを目的とする。

【0018】

【課題を解決するための手段】識別対象のDNAにユニークなプローブが選択できない場合には、以下の方針で識別する。

- ・複数のプローブを用いて、正しく識別できる確率を高ぐする。
- ・複数種類のDNAが同時に存在する可能性を考慮して、正しく識別できる確率を最大化するプローブの組を選別する。

これにより、識別精度を向上できる。

【0019】識別対象のDNAが後から加えられたときにには、以下の方針で新たなDNAに対応したDNAチップを製造する。

- ・最初にプローブを設計するときに、一種類のDNAに対

して複数の異なるプローブを用意しておく。新たに識別対象のDNAが加わり一部のプローブが使えなくなってしまい、残りのプローブを用いて識別できる。

【0020】また、既存のプローブでは識別できなくなった場合も、医学的・生物学的知識を用いて識別できるのであれば、プローブを追加しない。これにより、新たな細菌に対応したDNAチップを製造するためには、一部のDNAについてプローブを追加するだけで済み、検証実験を行ったりプローブ用DNA断片を用意したりするための時間的・金銭的コストが削減される。

【0021】本発明によるプローブ設計装置及びプローブ設計方法は、以下のようないくつかの特徴を有する。

(1) ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計装置において、塩基配列を含む識別対象ターゲットの情報を登録する識別対象ターゲット登録処理部と、登録された各ターゲットに対してその塩基配列の部分配列から各ターゲットにユニークな部分配列をプローブとして選択するプローブ選択処理部と、前記プローブ選択処理部においてユニークな部分配列が選択できなかったターゲットがある場合、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択する複数プローブ代用処理部とを含むことを特徴とするプローブ設計装置。

【0022】(2) ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計装置において、複数種類のターゲットを識別するための既存のプローブセットの情報を読み込む手段と、識別対象に追加する新たなターゲットの塩基配列を含む情報を読み込む識別対象ターゲット追加処理部と、前記新たなターゲットの追加によって前記既存のプローブセット内のプローブによる識別が困難になったターゲットを検出するプローブ再検討処理部と、前記新たなターゲットを識別するためのプローブ及び前記プローブ再検討部によって検出されたターゲットを識別するためのプローブを追加するプローブ追加処理部と、前記プローブ追加処理部においてユニークなプローブが選択できなかったターゲットがある場合、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択する複数プローブ代用処理部と、前記プローブ追加処理部及び/又は複数プローブ代用処理部による処理結果を反映してプローブセットを更新するプローブ更新処理部とを含むことを特徴とするプローブ設計装置。

【0023】(3) (1) 又は (2) のプローブ設計装置において、前記複数プローブ代用処理部は、複数のプローブとハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を一つの基準とし、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計装置。

(4) (1) 又は (2) のプローブ設計装置において、前記複数プローブ代用処理部は、混合感染・合併症の情報を用いて複数のプローブヒハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を評価し、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計装置。

【0024】(5) ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計方法において、識別対象であるターゲット集団の塩基配列を含むデータを読み込むステップと、読み込まれた各ターゲットの塩基配列の部分配列から各ターゲットにユニークな部分配列をプローブとして選択するステップと、前記ステップにおいてユニークな部分配列が選択できなかったターゲットがある場合、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択するステップとを含むことを特徴とするプローブ設計方法。

【0025】(6) ハイブリダイゼーションを介して識別対象である複数種類のターゲットを識別するためのプローブセットを設計するプローブ設計方法において、複数種類のターゲットを識別するための既存のプローブセットのデータを読み込むステップと、識別対象に追加する新たなターゲットの塩基配列を含むデータを読み込むステップと、前記新たなターゲットの追加によって前記既存のプローブセット内のプローブによる識別が困難になったターゲットを検出するステップと、前記新たなターゲットを識別するためのプローブ及び前記識別が困難になったターゲットを識別するためのプローブを追加するプローブ追加ステップと、前記プローブ追加ステップにおいてユニークなプローブが選択できなかったターゲットがあつたとき、複数のプローブの組み合わせを当該ターゲットに対する代用プローブとして選択する代用プローブ選択ステップと、前記プローブ追加ステップで追加されたプローブ及び/又は前記代用プローブ選択ステップによって選択されたプローブを反映してプローブセットを更新するステップとを含むことを特徴とするプローブ設計方法。

【0026】(7) (5) 又は (6) のプローブ設計方法において、前記代用プローブ選択ステップは、複数のプローブとハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を一つの基準とし、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計方法。

【0027】(8) (5) 又は (6) のプローブ設計装置において、前記代用プローブ選択ステップは、混合感染・合併症の情報を用いて複数のプローブヒハイブリダイズする複数種類のターゲットが試料中に同時に存在する確率を評価し、当該確率が低いプローブの組み合わせを代用プローブとして選択することを特徴とするプローブ設計方法。

【0028】(9) (5)～(8)のいずれか記載のプローブ設計方法をコンピュータに実行させるためのプログラム。

(10) (5)～(8)のいずれか記載のプローブ設計方法をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

#### 【0029】

【発明の実施の形態】以下では、細菌同定を例に挙げ、図面を参照して本発明の実施の形態を説明する。図4は、本発明のシステムの概略構成図である。このシステムは、細菌のDNA配列の情報や、細菌が検出される場所、合併症や混合感染などの医学的・生物学的情報を保存した細菌データベース400、既に設計したDNAチップの情報（識別対象としている細菌、それぞれの細菌に対するプローブの情報など）を保存したDNAチップデータベース401、設計したプローブの情報を出力するための表示装置402、識別対象に加える細菌や更新するDNAチップを選択するためのキーボード403及びマウスなどのポインティングデバイス404、プローブ選別を行う中央処理装置405、中央処理装置405での処理に必要なプログラムを格納するプログラムメモリ406、中央処理装置405での処理に必要なデータを格納するデータメモリ415から成る。

【0030】プログラムメモリ406は、新規にDNAチップを設計する際に識別対象細菌を登録する識別対象細菌登録処理部407、既存のDNAチップに対して識別対象の細菌を追加登録する識別対象細菌追加処理部408、新規にDNAチップを設計する際にプローブを選別するプローブ新規選択処理部409、既存のDNAチップに対して識別対象の細菌を追加する場合、既存のプローブと実験条件が揃うように、プローブを選び足すプローブ更新処理部410、設計したプローブ情報を表示するプローブ情報表示処理部411を備える。プローブ新規設計処理部409とプローブ更新処理部410は、識別対象のDNAにユニークなプローブが設計できなかった場合に複数のプローブで代用する複数プローブ代用処理部411を含む。プローブ更新処理部410は、新しく識別対象のDNAが追加されても既存のプローブで識別できるかどうかを判断するプローブ再検査処理部412、既存のプローブと実験条件が揃うように、プローブを選び足すプローブ追加処理部413を含む。データメモリ415は、細菌の医学的・生物学的情報のデータを保持する細菌データ416、及びDNAチップのデータを保持するDNAチップデータ417を備える。

【0031】図24は、識別対象細菌登録処理部407による処理画面（識別対象細菌登録画面）の例である。この処理画面は、新規にDNAチップを設計するための識別対象の細菌を登録する際に用いる。2400は、DNAチップの識別名である。2401は、細菌データベース400に登録されている細菌のリストである。ここか

ら、識別対象とする細菌を選択する。2402は、識別対象として選択した細菌を表示する列である。選択した細菌の欄にはレ印が表示される。2403は、細菌の名前（和名）である。2404は、細菌の名前（学名）である。2405は、細菌が検出される場所である。ここでは菌の和名と学名と検出される場所のみを表示しているが、病状やDNA配列データなどの情報を表示してもよい。2406は、選択した細菌の数であり、2402に表示されている列の個数と同じ値である。2407は、決定ボタンである。このボタンを押すと、現在選択している細菌が識別対象として確定される。

【0032】図25は、識別対象細菌追加処理部408による処理画面（識別対象細菌追加画面）の例である。この処理画面は、既存のDNAチップに対し識別対象の細菌を追加する際に用いる。2500は、DNAチップの識別名である。2501は、このDNAチップを設計した年月日である。2502は、このDNAチップを最後に更新した年月日である。設計・更新とも、ここでは年（西暦）と月と日のみ表示しているが、年に元号を用いたり年月日の他に時間を表示したりしても良い。また、更新履歴などの情報を表示しても良い。2503は、細菌データベースに登録されている細菌のリストであり、2401と同じ物である。ここから、識別対象に加える細菌を選択する。2504は、識別対象として追加した細菌を表示する列である。元から識別対象だった細菌の欄はレ印が表示された状態で固定されている。元は識別対象としていなかった細菌のうちで新たに識別対象に加える細菌を選択すると、新たにレ印が表示される。2505は、細菌の名前（和名）である。2506は、細菌の名前（学名）である。2507は、細菌が検出される場所である。ここでは菌の和名と学名と検出される場所のみを表示しているが、病状やDNA配列データなどの情報を表示してもよい。2508は、元から識別対象だった細菌の数であり、2504でレ印が表示された状態で固定されている細菌の数と同じ値である。2509は、新たに追加する細菌の数であり、2504に加わったレ印の数と同じ値である。2510は、決定ボタンである。このボタンを押すと、現在選択している細菌が識別対象として追加される。2511は、キャンセルボタンである。このボタンを押すと、識別対象としている細菌は変更されないことになる。

【0033】図5は、プローブ表示処理部414による処理画面（プローブ選別結果表示画面）の例である。この処理画面は、本システムが選別したプローブの情報を表示する際に用いる。500は、DNAチップの識別名である。501は、このDNAチップを設計した年月日である。502は、このDNAチップを最後に更新した年月日である。設計・更新とも、ここでは年（西暦）と月と日のみ表示しているが、年に元号を用いたり年月日の他に時間を表示したりしても良い。また、更新履歴などの情

報を表示しても良い。503は、このDNAチップが識別対象としている細菌の数である。504は、このDNAチップが識別対象としている細菌の一覧表である。表の行数は、503と同じ値である。マウス404やキーボード403を用いて、このリストから細菌を選択できる。ここでは菌の識別番号と和名と対応するプローブの数のみを表示しているが、学名やDNA配列データなどの情報を表示してもよい。505は、リスト504から選択された細菌に対応するプローブの一覧表である。ここではプローブの長さ、細菌のDNAにおける位置、反応温度のみを表示しているが、自身で略み合ってしまう確率や他の細菌のNAと誤って反応してしまう確率、他のプローブとの位置関係などを表示してもよい。

【0034】図4は、細菌データベース400から読み出して、データメモリ415の細菌データ416で保持する細菌の医学的・生物学的情報のデータ構造の例である。600は、細菌のID番号である。全ての細菌について一意である。この番号を用いて細菌を識別する。601は、和名である。602は、学名である。603は、DNA配列である。604は、この細菌が検出されるる場所である。605は、この細菌と合併症を起こしたり混合感染したりする可能性のある細菌のID番号600の一覧である。ここでは細菌IDが3の細菌との混合感染と、細菌IDが15及び24の細菌との混合感染の可能性があることを示す。605は冗長性のある記述になっている。すなわち、細菌ID3、15及び24の細菌の605には、それぞれ、(1と3)、(1と15と24)、(1と15と24)が含まれている。これらのはか、種や属、症状などの情報を保持しても良い。

【0035】図7は、DNAチップデータベース401から読み出して、データメモリ415のDNAチップデータ417で保持するDNAチップのデータ構造の例である。本システムでは、DNAチップのプローブの設計・更新にこのデータ構造を用いる。700は、このDNAチップの識別名である。図5のプローブ選別結果表示画面ではチップ識別名500にこの識別名を表示する。701および702は、このDNAチップの設計日および最終更新日である。プローブ選別結果表示画面では501と502に表示する。703は、識別対象としている細菌の数である。プローブ選別結果表示画面では503に表示する。704は、識別対象としている細菌の集合である。次に示すTargetData構造体のデータをリスト構造で保持し、先頭の要素へのポインタをここに格納する。リストの要素数は703の値と同じである。

【0036】705から708は、TargetData構造体のデータを示している。705は、細菌のID番号である。706は、この細菌を識別するために選別されたプローブの数である。707は、この細菌を識別するために選別されたプローブの集合である。次に示すProbeData構造体のデータをリスト構造で保持し、先頭の要素へのポ

インタをここに格納する。リストの要素数は706の値と同じである。708は、次のTargetData構造体を指すポインタである。最後の細菌だった場合は、ヌルポインタになる。

【0037】709から716は、ProbeData構造体のデータを示している。709は、プローブのDNA配列である。プローブ選別結果表示画面の505にこの情報を表示する。710は、プローブのDNA中ににおける開始位置で、細菌のDNA配列の先頭から数えて何塩基目であるかを示す。プローブ選別結果表示画面の505にこの情報を表示する。711は、プローブのハイブリダイゼーション温度である。プローブ選別結果表示画面の505にこの情報を表示する。712は、このプローブが自分自身で略み合ってしまう度合いである。プローブ選別結果表示画面の505にこの情報を表示しても良い。713は、このプローブが誤って他の細菌とハイブリダイゼーションしてしまう度合いである。プローブ選別結果表示画面の505にこの情報を表示しても良い。714および715は、このプローブのDNAチップ上の位置を示す。それぞれ、細胞Noと横Noの値を保持する。716は、次のProbeData構造体を指すポインタである。最後のプローブだった場合は、ヌルポインタになる。

【0038】図8は、本発明の概略処理フローである。以下、フローに沿って処理の流れを説明する。まず、キーボード403やマウス404などを用いて、DNAチップ上のプローブを新規に設計するのか、既存のDNAチップ上のプローブ情報を更新するのかをユーザーに選択させる(ステップ801)。ステップ801において、新規にDNAチップを設計するのであれば、まず、識別対象細菌登録処理部407に登録する処理を実行する。すなわち、キーボード403やマウス404などを用いて、識別対象の細菌の集合をユーザーに入力させ、対応するデータを細菌データベース400から読み込む(ステップ802)。ここで読み込むデータの構造は、図6で示したBacteriaDataである。ここでの画面イメージは、図24に示した通りである。選択された細菌の情報をデータメモリの細菌データ416及びデータメモリのDNAチップデータ417のTargetDataに登録し、次の要素へのポインタ708によってリスト構造でつなげる。次に、プローブデータ構造体409による処理を実行する。すなわち、これらのデータに基づき、DNAチップ上のプローブの選択を行う(ステップ803)。プローブ選択の処理の詳細は、後で説明する。

【0039】ステップ801において、既存のDNAチップ上のプローブ情報を更新するのであれば、まず、キーボード403やマウス404などを用いて、更新対象のDNAチップをユーザーに選択させ、対応するデータをDNAチップデータベース401から読み込み、データメモリのDNAチップデータ417に登録する(ステップ804)。ここで読み込むデータの構造は、図7で示したCh

ipDataである。また、選択したDNAチップが識別対象としている細菌のデータを細菌データベース4000から読み込む。ここで読み込むデータの構造は、図6で示したBacterialDataである。次に、識別対象細菌追加処理部408による処理を実行する。すなわち、キーボード403やマウス404などを用いて、新たに識別対象に加える細菌をユーザに選択させ、対応するデータを細菌データベース4000から読み込む(ステップ805)。ここで読み込むデータの構造は、図6で示したものである。ここで画面イメージは、図25に示した通りである。選択された細菌の情報をデータメモリのDNAチップデータ416及びデータメモリのDNAチップデータ417のtargetDataに登録し、次の要素へのポインタ708によってリスト構造を伸ばす。その後、プローブ更新処理部410による処理を実行する。すなわち、これらのデータに基づき、DNAチップ上のプローブ情報を更新する(ステップ806)。プローブ情報更新処理部414による処理を実行する。すなわち、選別結果を表示し、DNAチップデータベース401に保存する(ステップ807)。ここで画面イメージは、図5に示した通りである。

【0040】図9は、プローブ新規選択処理部409による処理、すなわち、図8における新規にDNAチップ上のプローブを選択する処理(ステップ803)の詳細フローである。以下、フローに沿って処理の流れを説明する。まず、それぞれの細菌に対して、その部分配列を列举し、プローブ候補の集合となる(ステップ900)。図10に示すように、その細菌のDNA配列に含まれる全ての部分配列を、プローブ候補とし、その後、 $T_m$ 値や自身自身で絡み合う度合い、ミスハイブリダイゼーションする度合いなどの条件を満たすプローブ候補を選別する(ステップ901～903)。図10では、プローブの長さは全て同じになっているが、さまざまな長さのプローブを設計しても構わない。その場合は、それぞれの長さの部分文字列を列举する。

【0041】全てのプローブ候補の $T_m$ 値が均一になるように、プローブ候補を選別する(ステップ901)。 $T_m$ 値とは、DNAの二本鎖が一本鎖になるとき(またはその逆)の温度のことであり、DNA配列中に含まれるG(グアニン)とC(シトシン)の割合(GC含量)によって決まる。DNAチップ上のプローブは、全て同じ温度で実験されるので、 $T_m$ 値が均一になっていることが望ましい。

【0042】図11は、 $T_m$ 値が均一になるようにプローブを選別する状況を示した図である。細菌Pのプローブ候補が三つ(プローブ候補1、2、3)あり、それぞれGC含量は2、1、3である。また、細菌Qのプローブ候補が二つ(プローブ候補1、2)あり、それぞれ、GC含量は3、4である。このような場合、細菌Pのプローブ候補①と細菌Qのプローブ候補①を選んだ場合は、GC含量が3になり、 $T_m$ 値が均一になる。 $T_m$ 値のばらつきは5

°C以内程度に抑えるのが好ましい。

【0043】それぞれの細菌について、プローブ候補から自分自身で絡み合わないものを選別する(ステップ902)。図12に示すように、プローブが自分自身で絡み合ってしまうと、試料中のDNAを捕獲することができない。従って、検出の感度を上げるためにには、絡み合いにくいプローブを選別する必要がある。プローブが絡み合いやすいかどうかの判断基準としては、スタッキング構造・バルジループ構造・インターナルループ構造・ヘアピン構造・分岐ループ構造などに対してNearest-Neighbor法を用いて自由エネルギーを計算する手法が知られている。

【0044】それぞれの細菌について、誤って他の細菌とハイブリダイゼーションにくいプローブを選別する(ステップ903)。図13に示すように、プローブと試料中のDNAとが類似していると部分的にハイブリダイゼーションしてしまうことがあり、これはミスハイブリダイゼーションと呼ばれている。図13の左の例ではプローブの中ほどでTとGになっており、右の例ではプローブの先端でTとTになっているが、その他の全てでT、Gとの組合せになっている。このようにミスハイブリダイゼーションしてしまうと、プローブに対応しない細菌が誤って検出される。

【0045】それぞれの細菌について、プローブ候補のうちユニークなものを選ぶ。ユニークなプローブ候補が存在しない細菌に対しては、複数プローブ代用処理部411を実行し、複数のプローブ候補で代用する。複数のプローブで代用する処理の詳細は後で説明する(ステップ904)。

【0046】最後に、ステップ900～904で選ばれたプローブ候補をプローブとする(ステップ905)。すなわち、それぞれの細菌に対応するプローブの情報を、プローブの情報をメンバ709～715に登録し、次の要素へのポインタ716によってリスト構造でつなげていく。ここではプローブが満たすべき条件を、反応温度、自分自身で絡み合う度合い、他の細菌のDNAに誤ってハイブリダイゼーションする度合い、ユニークであること、の順に考慮したが、これらの順序は変更しても構わない。

【0047】図14は、プローブ更新処理部410による処理。すなわち、図8における既存のDNAチップ上のプローブ情報を更新する処理(ステップ806)の詳細フローである。以下、フローに沿って処理の流れを説明する。まず、プローブ追加処理部413による処理を実行し、新しく加えられた細菌Xを識別するために追加するプローブを選び、ChipDataに登録する(ステップ1400)。追加するプローブを選ぶ処理については、後で詳しく説明する。次に、プローブ再検討処理部412による処理を実行し、今まで識別対象としていた細菌のうち、識別できなくなった細菌があるか調べる(ステップ

1401)。この処理についても、後で詳しく説明する。今まで識別対象としていた細菌のうち、細菌Xが識別対象に加わったために識別できなくなってしまうもの(細菌Yと呼ぶ)が存在する場合がある。このような細菌Yが存在する場合は、プローブ追加処理部413による処理を実行して細菌Yを識別するためのプローブを追加し、プローブのリスト704を更新する(ステップ1402、1403)。また、ステップ1402において識別できない細菌が一つもない場合は、処理を終了する。

【0048】図15は、プローブ追加処理部413による処理、すなわち、図14における新しく加えられた細菌Xを識別するために追加するプローブを選択する処理(ステップ1400)の詳細フローである。細菌Yを識別するためのプローブを追加する処理(ステップ1403)も同様である。以下、フローに沿って処理の流れを説明する。

【0049】まず、細菌Xの部分配列を列挙し、プローブ候補の集合とする(ステップ1505)。すでにいくつかプローブを設計してある細菌に対して追加のプローブを選択するときは、候補の集合から既存のプローブを除いておく。次に、プローブ候補のうち、ほかのプローブとTmが近いものを選ぶ(ステップ1501)。

【0050】図16は、新しく識別対象に加わった細菌Xに対して、ほかのプローブとTmが近いプローブ候補を選択状況を示した図である。細菌P、Qを識別対象として設計されたDNAチップ、新たに識別対象として細菌Xが加わったとする。細菌P、QのプローブのGC含量は1Xは2である。したがって、この場合細菌Xのプローブについてもこの範囲のGC含量を持つもの(細菌Xのプローブ候補1、3)を用いければ、DNAチップ全体のTm値を均一にすることができる、実験精度を高められる。その後、プローブ候補のうち、自分自身で絡み合わないものを選ぶ(ステップ1502)。この処理は、ステップ902における処理と同様である。

【0051】他の細菌のDNAと誤ってハイブリダイゼーションする確率が低い候補を選択(ステップ1503)。この処理は、ステップ903における処理と同様である。プローブ候補のうち、細菌Xにユニークなものががあれば、それを細菌Xのプローブとする(ステップ1504)。ユニークなプローブ候補が存在しなかったら、複数のプローブで代用する。複数のプローブで代用する処理の詳細は後で説明する。

【0052】このようにして選ばれたプローブ候補を細菌Xのプローブとし、ProbeDataに登録する。すなわち、プローブの情報をスマバ709~715に登録し、次の要素へのポイント716によってリスト構造でつなげていく(ステップ1505)。ここでも図9同様、プローブが満たすべき条件を、反応温度、自分自身で絡み合う度合い、他の細菌のDNAと誤ってハイブリダイゼーションする

する度合い、ユニークであること、の順に考慮したが、これらの順序は変更しても構わない。

【0053】図17は、プローブ再検討処理部412による処理。すなわち、図14における識別できない細菌があるか調べる処理(ステップ1401)の詳細フローである。フローに示す処理を、それぞれの細菌について行う。以下、フローに沿って処理の流れを説明する。

【0054】まず、細菌Yのプローブのうち、細菌Xとも反応してしまうものをプローブのリスト707から除く(ステップ1700)。なぜなら、後から加えられた細菌のDNA配列が、プローブの配列を部分配列として含む場合があるからである。図18は、細菌P、Q、Yを識別対象としてプローブを設計しておいたところへ、新たに識別対象の細菌Xとして細菌Xが追加された状況を示している。細菌のDNA配列は、細菌Yのプローブを部分配列として含むため、細菌Yのプローブは細菌X、Yの両方と反応する。細菌Xが追加されたことにより、細菌Yのプローブを用いた細菌Yの存在を正しく判断できなくなってしまった。

【0055】次に、細菌Yのプローブのうち、細菌Xとミスハイブリダイゼーションをしてしまうものをプローブのリスト707から除く(ステップ1701)。なぜなら、新しく識別対象に加わった細菌Xと細菌Yのプローブが高い確率でミスハイブリダイゼーションする場合があるからである。図19は、細菌P、Q、Yを識別対象としてプローブを設計しておいたところへ、新たに識別対象の細菌Xが追加された状況を示している。細菌Yのプローブは、細菌P、QのDNA配列と類似度が低いので、細菌Xが加わるまではミスハイブリダイゼーションが起る可能性は低かった。しかし、細菌XのDNA配列は細菌Yのプローブと類似度が高いので細菌XとXを誤って判断してしまう可能性が高い。細菌Xが追加されたことにより、細菌Yのプローブは最適ではなくなってしまった。

【0056】ステップ1700と1701でプローブを除いた結果、プローブのリストが空になったかどうか調べる(ステップ1702)。空になってしまったのであれば、細菌Xと識別対象に加わった細菌Yを識別することができないので、ステップ1403で説明したようにプローブを追加する必要がある。空にならなかつたのであれば、残ったプローブで細菌Yを識別できるので、プローブを追加する必要はない。

【0057】図20は、複数プローブ代用処理部411による処理。すなわち、図9および図15における識別対象の細菌Xにユニークなプローブが設計できない場合に複数のプローブで代用する処理(ステップ904および1504)の詳細フローである。図21は、細菌ZのDNA配列が細菌B1、B2、B3、B4と共通であるため、ユニークな部分配列を選択できない様子を示したものである。以下では簡単のため、細菌B1、B2、B3、B4のそれぞれに

についてはユニークなプローブを設計できるものとして、処理の流れを説明する。

【0058】まず、細菌Zの部分配列を全て列挙し、プローブ候補とする（ステップ2000）。図21では説明を簡単にするため、P1、P2、P3の三つのみをプローブの候補とするが、実際には細菌Zの全ての部分配列をプローブ候補とする。次に、プローブ候補のそれぞれについて、細菌Z以外に反応してしまう細菌がどれだけあるか調べる（ステップ2001）。図21では、プローブ候補1は細菌B1ともハイブリダイゼーションし、プローブ候補2は細菌B2ともハイブリダイゼーションし、プローブ候補3は細菌B3、B4ともハイブリダイゼーションする。

【0059】その後、以下の二つの方針で正しく識別できる確率を高くする（ステップ2002）。第一の方針は、複数のプローブを用いることである。まず、P1を細菌Zのプローブとして用いる場合を考える。図22に示すように、細菌B1が存在すると細菌Zに間わらずP1からシグナルが観測される。したがって、細菌B1が存在すると（細菌B2の存在には間わらず）、細菌B2が存在するのかどうかプローブのシグナルだけからは判断できない。次に、P1とP2を細菌Zのプローブとして用いる場合を考える。図22に示すように、細菌B1とB2が両方存在すると細菌Zの存在に間わらずP1とP2の両方からシグナルが観測される。したがって、細菌B1と細菌B2が両方存在すると、細菌Zが存在するのかどうか判断できない。「細菌B1とB2が両方とも存在する確率」は「細菌B1が存在する確率（細菌B2は存在してもしなくても良い）」より小さいで、P1とP2の両方を用いることにより、細菌Zが存在するかどうか判断できない確率を減少できる。プローブP1、P2、P3の全てを用いるのであれば、さらにこの確率を減少できる。しかし、プローブの数を増やすとそれだけコストが増えてしまう。そこで、以下で説明する第二の方針にしたがってプローブ候補を絞り込む。

【0060】第二の方針は、より少ないプローブで「細菌Zが存在するのかどうか判断できない確率」を下げる事である。例えば、プローブ候補を二つ選ぶ場合、「P1とP2」「P2とP3」「P1とP3」という選択肢のそれぞれについて「細菌Zが存在するのかどうか判断できない確率」を計算し、これが最小となるプローブ候補の組合せを選ぶ。図22で示した通り、P1とP2を用いる場合は、細菌B1とB2が両方とも存在すると、細菌Zが存在するのかどうか判断できない。図23に示すように「P2とP3」「P1とP3」についても、どのような場合に判断できないか求める。次に、「細菌B1と細菌B2が同時に試料中に存在する可能性」「細菌B2が存在し、かつ、細菌B3またはB4が存在する可能性」「細菌B1が存在し、かつ、細菌B3またはB4が存在する可能性」を考える。下の例に挙げるように、合併症や混合感染をする菌

が知られている。

【0061】例1：インフルエンザは細菌性肺炎との合併症を起こしやすい。（加古川市加古郡医師会感染症情報提供システム HYPERLINK "http://www.kakogawa.or.jp/kakomed/memo20.htm" http://www.kakogawa.or.jp/kakomed/memo20.htmより）

例2：MRSA（メチシリン耐性黄色ブドウ球菌）は線維菌と同時に存在する確率が高（引佐赤十字病院のホームページ HYPERLINK "http://www.habi.ne.jp/rc-inasa/topics.html" http://www.habi.ne.jp/rc-inasa/topics.htmlより）

【0062】例3：パンコマイシン耐性腸球菌と線維菌や大腸菌は混合感染しやすい。（感染症発生動向調査週報・感染症の話 2000年第20週（5月15日～5月21日）、第21週（5月22日～5月28日） HYPERLINK "http://idsc.nih.go.jp/kansen/k00-g30/k00#201/k00#201.htm" http://idsc.nih.go.jp/kansen/k00-g30/k00#201/k00#201.htmより）

【0063】例4：クラミジア肺炎は肺炎マイコプラスマや肺炎球菌との混合感染が多い。（感染症発生動向調査週報・感染症の話 1999年第45週（11月8日～14日） HYPERLINK "http://idsc.nih.go.jp/kansen/k99-g52/k99#45.htm" http://idsc.nih.go.jp/kansen/k99-g52/k99#45.htmより）

例5：ジグルジア症は赤痢菌や膿原大腸菌や赤痢アメバと混合感染する（感染症発生動向調査週報・感染症の話 2000年第12週（3月20日～3月26日） HYPERLINK "http://idsc.nih.go.jp/kansen/k00-g15/k00#12/k00#12.htm" http://idsc.nih.go.jp/kansen/k00-g15/k00#12/k00#12.htmより）

【0064】例6：赤痢アメバ感染症は他の性感染症（梅毒、HIV感染症、B型肝炎、性器ヘルペスなど）を合併していることがある（感染症発生動向調査週報・感染症の話 2000年第29週（7月17日～7月23日） HYPERLINK "http://idsc.nih.go.jp/kansen/k00-g30/k00#29/k00#29.htm" http://idsc.nih.go.jp/kansen/k00-g30/k00#29/k00#29.htmより）

【0065】例7：エイズはカリニ肺炎や口腔内カンジダやトキソプラズマ症や帶状疱疹や非定型抗酸菌症や赤痢アメバや抗真菌薬やサイトメガロウイルス感染症と合併する（感染症発生動向調査週報・感染症の話 2000年第23週（6月5日～6月11日） HYPERLINK "http://idsc.nih.go.jp/kansen/k00-g30/k00#23/k00#23.htm" http://idsc.nih.go.jp/kansen/k00-g30/k00#23/k00#23.htmより）

【0066】例8：グラム陰性桿菌とカンジダ菌は混合感染をする。（石川 博人 et al.：カンジダ血症は悪性腫瘍に対する腎切削後細胞免疫を低下させる：第98回日本外科学会総会 演題番号P-114 HYPERLINK "http://www.jss.h.u-tokyo.ac.jp/JSS/Abstracts/P-114.htm" http://www.jss.h.u-tokyo.ac.jp/JSS/Abstracts/P-114.htmより）

4.htm、山口 英世：真菌感染症の問題点について：第24回 日本熱傷学会総会・学術集会 教育講演 HYPERLINK "http://www.med.teikyo-u.ac.jp/tccc/burn24/index.htm" http://www.med.teikyo-u.ac.jp/tccc/burn24/index.htmより）

【0067】例9：ジフテリア菌は連鎖球菌やブドウ球菌と混合感染をする。（ジフテリア予防対策マニュアル：国立感染症研究所 細菌・血液製剤部 衛生微生物技術協議会レファレンス委員会 ジフテリア小委員会 HYPERLINK "http://idsc.nih.go.jp/others/diphtheria/jifuteria.html" http://idsc.nih.go.jp/others/diphtheria/jifuteria.htmlにより）

例10：チフス菌と赤痢菌やサルモネラやランブル鞭毛虫は混合感染をする。（病原微生物検出情報Vol.17 No.12:IASR HYPERLINK "http://idsc.nih.go.jp/iasr/index-j.html" http://idsc.nih.go.jp/iasr/index-j.htmlより）

【0068】上記のような、合併症を起こしやすい菌の組合せは試料中に同時に存在する可能性が高い。図23では細菌1と細菌3又は3&4は合併症を起こしやすく同時に存在する可能性が高いので、プローブ候補の「P1とP3」という組合せは、「P1とP2」「P2とP3」に比べて「細菌Zが存在するのかどうか判断できない確率」が高いことを示している。このような合併症・混合感染の情報を細菌データベース400のデータ605から得て、「細菌Zが存在するのかどうか判断できない確率」を低く抑えるように、プローブ候補の組を選んでいく。

【0069】以上の説明では例として細菌同定を挙げたが、試料中に混入したその他の生物種・DNAの識別においても同様である。また、複数種類の識別対象のDNAが試料中に同時に存在する可能性を検討するための判断材料として、細菌の合併症や混合感染のみを挙げたが、生育条件や同じかどうか、捕食関係、競合関係、共生関係にあるかどうかなどの医学的・生物学的知識を用いて判断することも考えられる。

【0070】【発明の効果】以上説明したように、本発明によれば、新たな細菌に対応したDNAチップを製造するための時間的・金銭的コストを削減することができる。また、ユニークなプローブが設計できない場合でも、高い確率で識別することが可能なDNAチップを製造できる。

【図面の簡単な説明】

【図1】DNAチップを利用した細菌同定の模式図。

【図2】DNAチップ上に配置するプローブの選別過程を示した図。

【図3】識別対象の細菌のDNAにユニークなプローブを

設計できない例を示した図。

【図4】本発明のDNAチップシステムの構成を示したブロック図。

【図5】表示装置の表示画面の一例を示す説明図。

【図6】細菌データのデータ構造を示した図。

【図7】DNAチップデータのデータ構造を示した図。

【図8】本発明の概略処理フローを示す図。

【図9】プローブ選別処理の詳細フロー図。

【図10】プローブ候補の集合の変化を説明する図。

【図11】反応温度を均一にする処理の説明図。

【図12】プローブが自分で絡み合ってしまうことの説明図。

【図13】プローブと試料中のDNAが部分的にハイブリダイゼーションする例を示す図。

【図14】プローブを更新する処理の詳細フロー図。

【図15】プローブ追加処理の詳細フロー図。

【図16】後から識別対象に加えられた細菌に対し、他のプローブと反応温度が同じになるプローブを選択処理を示した図。

【図17】識別できない細菌があるか調べる処理の詳細フロー図。

【図18】後から識別対象の細菌が加えられたことにより、プローブがユニークでなくなることを示した図。

【図19】後から識別対象の細菌が加えられたことにより、ミスハイブリダイゼーションの確率が高くなることを示した図。

【図20】識別対象の細菌にユニークなプローブが設計できない場合に、複数のプローブで代用する処理の詳細フロー図。

【図21】識別対象の細菌のDNAにユニークなプローブを設計できない場合に、複数のプローブ候補を選択することの説明図。

【図22】プローブ候補の選び方と、どのような場合に識別できないかを示した図。

【図23】プローブ候補の選び方によって、識別できない可能性が変わることを示した図。

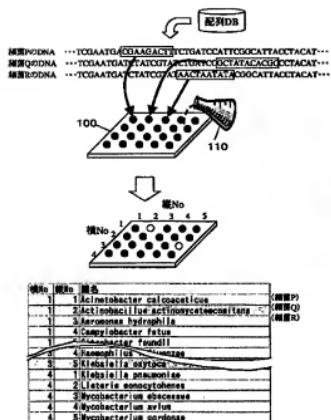
【図24】新規設計するDNAチップに対して識別対象の細菌を登録する処理における、表示装置の表示画面の一例を示す説明図。

【図25】既存のDNAチップに対して識別対象の細菌を追加登録する処理における、表示装置の表示画面の一例を示す説明図。

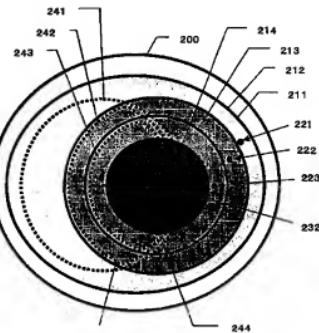
【符号の説明】

100…DNAチップ、110…試料、400…細菌データベース、401…DNAチップデータベース、405…中央処理装置、406…プログラムメモリ、415…データメモリ

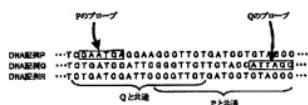
【図1】



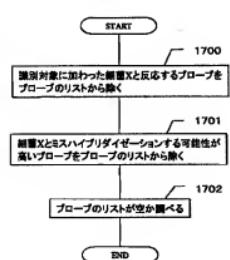
【図2】



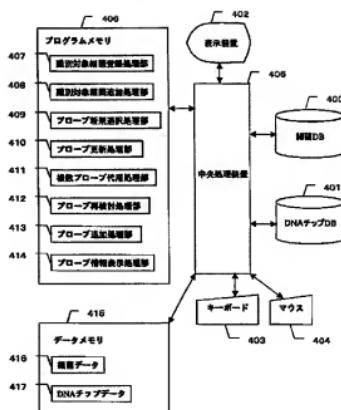
【図3】



【図17】



【図4】



〔图5〕

番号	部屋番号	フロア番号	部屋番号	フロア番号	部屋番号	フロア番号
1	大間屋	B	1	TTAATAGTAA ...	102	最適温度
2	青色アブリ屋	11	2	TTTAACTAGT ...	201	最適温度
3	マルセイウム屋	12	3	TTCTTGAGTAA ...	010	最適温度
4	ナフス屋	7	4	CTCTTGGGAA ...	002	最適温度
	⋮			⋮		
96			11			

【图6】

[图8]

【四七】

CtgData		PdgData	
番号	内容	番号	内容
700	メンバ名 チップ配列名	700	ID MC011818
701	検出日	701	2009年1月16日
702	最終更新日	702	2009年1月16日
703	チップの数	703	35
704	測定の場合は「ループ」	704	ループの場合は「ループ」
TupData x		PdgData	
705	メンバ名 検出ID	705	メンバ名 チップ配列
706	プローブの数	706	1
707	チップの場合は「ループ」	707	チップの場合は「ループ」
708	次の測定へのポインター	708	次の測定へのポインター
測定ID		PdgData	
709	メンバ名 プローブの数	709	メンバ名 チップ配列
710	プローブの場合は「ループ」	710	1
711	チップの場合は「ループ」	711	チップの場合は「ループ」
712	次の測定へのポインター	712	次の測定へのポインター
測定ID		PdgData	
713	メンバ名 プローブの数	713	メンバ名 チップ配列
714	プローブの場合は「ループ」	714	1
715	チップの場合は「ループ」	715	チップの場合は「ループ」
716	次の測定へのポインター	716	次の測定へのポインター

[図1.1]

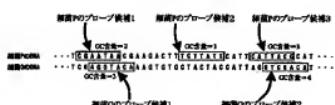
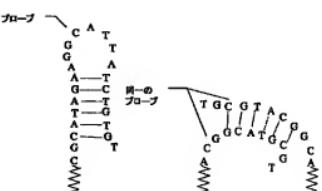
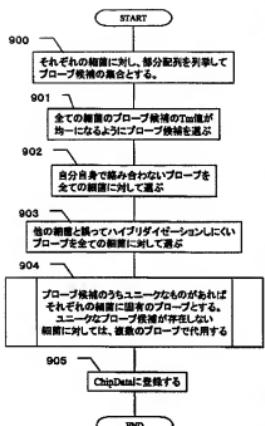


图127



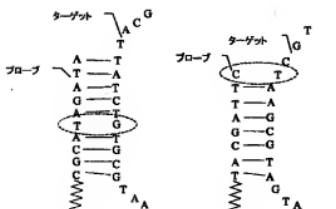
〔図9〕



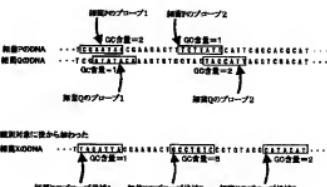
【图10】



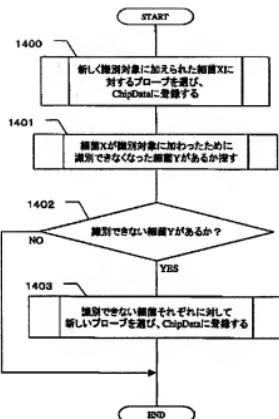
【図13】



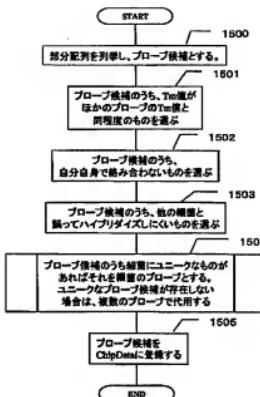
【图16】



〔図14〕



〔図15〕



[图18]

【图19】

新規DNA ... TCGATTA~~TCGATTA~~TCTTAATGCTTCCGATTCACAT  
既存DNA ... TCGATTA~~TCTTAATGCTTCCGATTCACAT~~TCGATTA  
既存RNA ... TCGATTA~~TCTTAATGCTTCCGATTCACAT~~TCGATTA

新規Qのプローブ  
既存Qのプローブ  
既存Yのプローブ

既存Yに他の塩基が混入された  
既存RNA ... TCGATTA~~TCGATTA~~TCTTAATGCTTCCGATTCACAT  
既存Yのプローブと既存Y/既存RNA

【圖2-1-1】

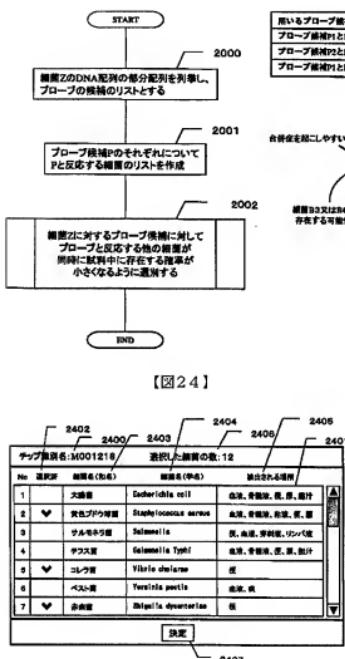
[图22]

プローブ候補P1と、候補2のプローブとして用いる場合			
候補の存在		プローブのレダナル	
候補P1	候補P2	候補P1のプローブ	プローブ候補P1
○	○	○	○
×	○	×	○
○	×	○	○

卷二：民族与区域：蒙古族与二胡音乐研究

周波数の存在			プローブのシグナル	
周波数1	周波数2	周波数3	プローブ候補1	プローブ候補2
○	○	○	○	○
×	○	○	○	○
○	×	○	○	○
×	×	○	○	○
○	○	×	○	○
×	○	○	×	○
○	×	×	○	×
×	×	×	×	×

【図20】

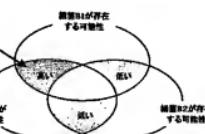


【図24】

2402	2400	2403	2404	2405	2401
サンプル番号: 13001218 選択した菌数の数: 12					
No. 菌名 菌名(学名) 菌名(学名) 菌名(学名) 菌名(学名) 菌名(学名)					
1 大腸菌 <i>Escherichia coli</i> 乳酸、キ酸、乳、酸、醣					
2	黒色グラム菌 <i>Staphylococcus aureus</i>	乳酸、キ酸、乳、酸、醣			
3	サルモネラ菌 <i>Salmonella</i>	乳、乳酸、キ酸、ソルビ			
4	サルモ菌 <i>Salmonella Typhi</i>	乳酸、キ酸、乳、酸、醣			
5	コラバ菌 <i>Vibrio cholerae</i>	乳			
6	ペスト菌 <i>Yersinia pestis</i>	乳酸、醣			
7	赤痢菌 <i>Shigella dysenteriae</i>	乳			
決定					
2407					

【図23】

用いるプローブ種類	判断できない場合
プローブ種類P1とP2	菌種3が菌種2が両方とも存在
プローブ種類P2とP3	菌種2が存在し、かつ、菌種3又は菌種4が存在
プローブ種類P1とP3	菌種1が存在し、かつ、菌種3又は菌種4が存在



【図25】

2504	2505	2506	2507	2503
チップ番号: NO0118	測定日: 2000年12月1日	最終更新日: 2001年4月11日		
選択された菌種の数: 12				
選択された菌種の数: 12				
No.	菌名	菌名(学名)	菌名(学名)	菌名(学名)
1	大腸菌	<i>Escherichia coli</i>	乳酸、キ酸、乳、酸、醣	
2	黒色グラム菌	<i>Staphylococcus aureus</i>	乳酸、キ酸、乳、酸、醣	
3	サルモネラ菌	<i>Salmonella</i>	乳、乳酸、キ酸、ソルビ	
4	サルモ菌	<i>Salmonella Typhi</i>	乳酸、キ酸、乳、酸、醣	
5	コラバ菌	<i>Vibrio cholerae</i>	乳	
6	ペスト菌	<i>Yersinia pestis</i>	乳酸、醣	
7	赤痢菌	<i>Shigella dysenteriae</i>	乳	
キャンセル 決定				
2511 2510				

フロントページの続き

(72)発明者 中重 充

神奈川県横浜市中区尾上町6丁目81番地  
日立ソフトウェアエンジニアリング株式会  
社内

(72)発明者 野崎 康行

神奈川県横浜市中区尾上町6丁目81番地  
日立ソフトウェアエンジニアリング株式会  
社内

(72)発明者 上野 韶吾

神奈川県横浜市中区尾上町6丁目81番地  
日立ソフトウェアエンジニアリング株式会  
社内

(72)発明者 田村 卓郎

神奈川県横浜市中区尾上町6丁目81番地  
日立ソフトウェアエンジニアリング株式会  
社内

F ターム(参考) 4B024 AA19 AA20 CA01 CA09 HA19

4B029 AA07 AA22 BB20 CC08 FA12

4B063 QA13 QQ42 QR32 QR55 QS34